BMC Genomics

DATABASE

Open Access

JOB: Japan Omics Browser provides integrative visualization of multi-omics data

Yugo Takahashi[1], Qingbo S. Wang[2,3*], Takanori Hasegawa[4], Ho Namkoong[5], Fumitaka Inoue[6], Koichi Fukunaga[7], Seiya Imoto[8], Satoru Miyano[4], Yukinori Okada[2,9,10,11*] and Japan COVID-19 Task Force

Abstract

We present the Japan Omics Browser (JOB), which enables integrative analysis of human omics at different layers. JOB offers visualization of per-variant regulatory effects in the human blood at mRNA and protein level distinctively, quantified from statistical fine-mapping of mRNA-expression quantitative loci (eQTL) and protein QTLs (pQTLs) in 1,405 Japanese, together with fine-mapping results of 94 complex traits in UK Biobank. In addition, JOB shows per-tissue regulatory effect prediction score (EMS), trained via multi-task learning. Furthermore, validation scores from Massively Parallel Reporter Assay (MPRA) in two cell types are available for over 10,000 variants. JOB is publicly available at https://japan-omics.jp/.

Keywords Multi-Omics, Integrative analysis, Fine-mapping, eQTL, pQTL, Regulatory effects, Machine-learning, Massively parallel reporter assay, Web tool, Database

*Correspondence:
Qingbo S. Wang
qingbow@m.u-tokyo.ac.jp
Yukinori Okada
yuki-okada@m.u-tokyo.ac.jp
[1] Faculty of Medicine, Osaka University, Suita, Japan
[2] Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Bunkyo, Tokyo, Japan
[3] Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan
[4] M&D Data Science Center, Institute of Integrated Research, Institute of Science Tokyo, Tokyo, Japan
[5] Department of Infectious Diseases, Keio University School of Medicine, Tokyo, Japan
[6] Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan
[7] Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan
[8] Division of Health Medical Intelligence, Human Genome Center, the Institute of Medical Science, the University of Tokyo, Tokyo, Japan
[9] Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
[10] Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan
[11] Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Japan

Background

In recent years, a wealth of human genome, transcriptome, proteome, and other omics data has become available, providing valuable resources for researchers and healthcare professionals to investigate genotype–phenotype relationships. These resources include large-scale genomic cohort studies, such as the 1000 Genomes Project and UK Biobank [1, 2], as well as databases such as NCBI GEO that is more focused on collecting transcriptomics data from individual high-throughput sequencing [3]. Efforts to catalog and publicly share multi-omics data are expanding, including those for gene expression patterns at the single-cell level, proteomics and other molecular trait omics data [4–6]. While these multi-omics datasets have provided insights into disease etiology, risk factors, and therapeutic approaches [7–9], handling such large-scale multimodal data poses challenges, particularly for clinical scientists with relatively limited set of skills to process high-dimensional data.

To maximize the user experience for researchers with diverse professional backgrounds, several web tools and software programs have been developed to visualize

and explore specific subsets of datasets. For instance, the UCSC Genome Browser allows for the exploration of annotated genome assemblies, while the gnomAD browser compiles genetic variant frequencies across diverse ethnic populations [10, 11]. The Biobank Japan pheweb allows for visualization of genetic associations of a variant of interest to a large number of traits [12]. The GTEx project provides a public resource of mRNA expression levels in various human tissues, facilitating comparative investigations into genetic variants and regulatory mechanisms [13]. Artificial intelligence techniques, such as deep learning, have demonstrated effectiveness in predicting the impact of variants on gene expression profiles [14], where user friendly visualization tools have also been developed [15].

Although such interactive browser style databases have expanded the usage of different layers of diverse omics data, the datasets are often scattered across individual databases, hindering comprehensive analysis in the face of exponential data growth. While databases aiming to integrate multiple layers of omics data with user-friendly interface exist [16], the major platform is not intended for optimal visualization of specific genomic variants or locus. In addition, existing databases predominantly focus on European populations especially when focused to molecular traits, resulting in limited representation of East Asian genomic variants and gene expression profiles [17, 18]. To bridge these gaps, it is essential to collect and share data on variants and expression patterns specific to the East Asian population and to allow visualization of multiple layers of variant effect in a single platform with clarity, enabling the analysis of interethnic genetic variations (Table 1).

Here, we have developed the Japan Omics Browser (JOB), an intuitive and public platform designed to empower researchers, including non-experts, to visually explore omics data primarily derived from the Japanese population (Fig. 1). JOB builds upon previously published datasets, including pQTL and eQTL studies conducted by the Japan COVID-19 Task Force (JCTF) [20, 21]. Additionally, JOB incorporates a machine learning-based score, which predicts the regulatory effects of variants on nearby gene expression in 49 tissues (Expression Modifier Score =EMS) [22]. Furthermore, the gene regulatory effects of these variants are functionally confirmed through the use of Massively Parallel Reporter Assay (MPRA) in two cell types (HepG2 and K562) for over 10,000 variants. We present multiple cases where, utilizing JOB, researchers can perform comprehensive omics analyses, gain valuable insights into disease etiology and discover potential biomarkers. JOB is publicly available at https://japan-omics.jp/.

Results

Overview of the Japan Omics Browser (JOB)

JOB is an integrated web application implemented on the Google Cloud Platform. To display qualitative and quantitative information for genes and variants in the JOB dataset, we have developed a scalable browser framework (Supplementary Material 1) allowing for seamless browsing on desktop, smartphones or any other devices (see Methods). On the search page, users can enter keywords such as position-based variant ID (chr:pos:ref:alt in hg38), rs ID (from dbSNP; https://www.ncbi.nlm.nih.gov/snp/), Ensembl gene ID, or canonical Gene name from HUGO (https://www.genenames.org) to navigate to specific pages (Fig. 2a).

When a variant ID or rs ID is entered in the search page, it redirects to the variant page (Fig. 2b). On the variant page, a table displays a list of genes located within the cis-window of variant position, defined by the distance to transcription starting site (TSS) smaller than 1 Mb, where the columns are relevant omics information. Users can find the description of each column by putting the cursor on the column, and users can sort the table based on each score. The omics features available in the table include:

1. (Marginal) effect sizes (beta), standard error of the effect sizes, minor allele frequency (MAF), *p*-value, posterior inclusion probability (PIP) from statistical fine-mapping of mRNA expression quantitative loci (eQTL) effect in two tools (SuSiE [23] and FINEMAP [24]).

Table 1 Comparison of JOB and other major platforms for multi-omics data visualization

	GWAS	eQTL	pQTL	Regulatory effect prediction	MPRA	East Asian Focus
JOB	o	o	o	o	o	o
GTEx [13]	x	o	x	x	x	x
UCSC Genome Browser [11]	x	x	x	o	x	x
Open Target [19]	o	o	o	x	x	x
BBJ-Pheweb [12]	o	x	x	x	x	o

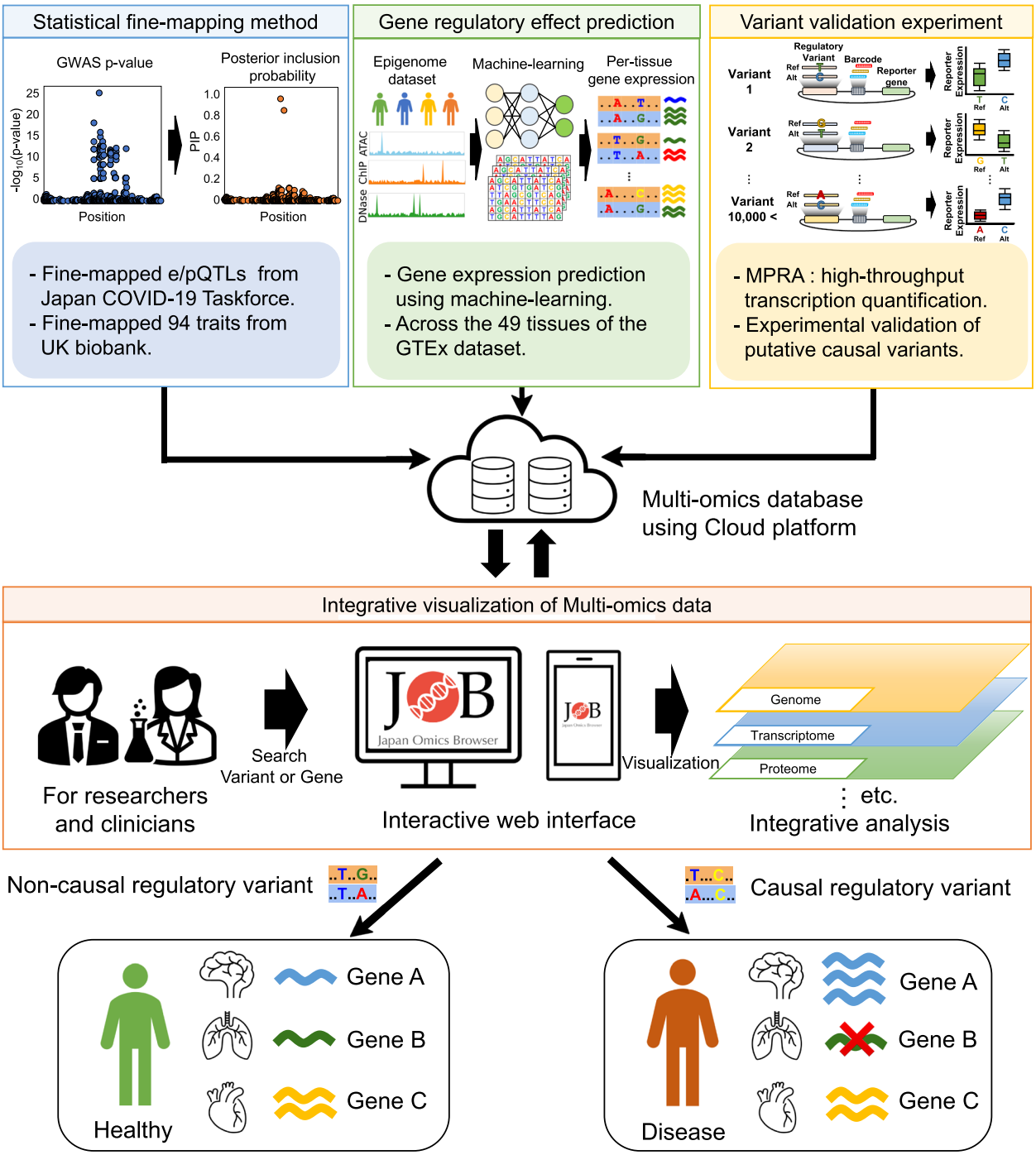

Fig. 1 Overview of the Japan Omics Browser (JOB). JOB is a database browser that integrates e/pQTL data from the Japan COVID-19 Taskforce, phenotypedata from the UK Biobank, and machine learning-based prediction of gene regulatory effect variants (EMS [22]). Additionally, it incorporates experimental validation scores from massively parallel reporterassays (MPRAs) for a subset of over 10,000 variants. Hosted on the Google Cloud Platform, JOB providesan interactive interface accessible on various devices. Users can search for variants/genes of interest andconduct integrated analyses to investigate their effects on mRNA/protein expression and complex traits

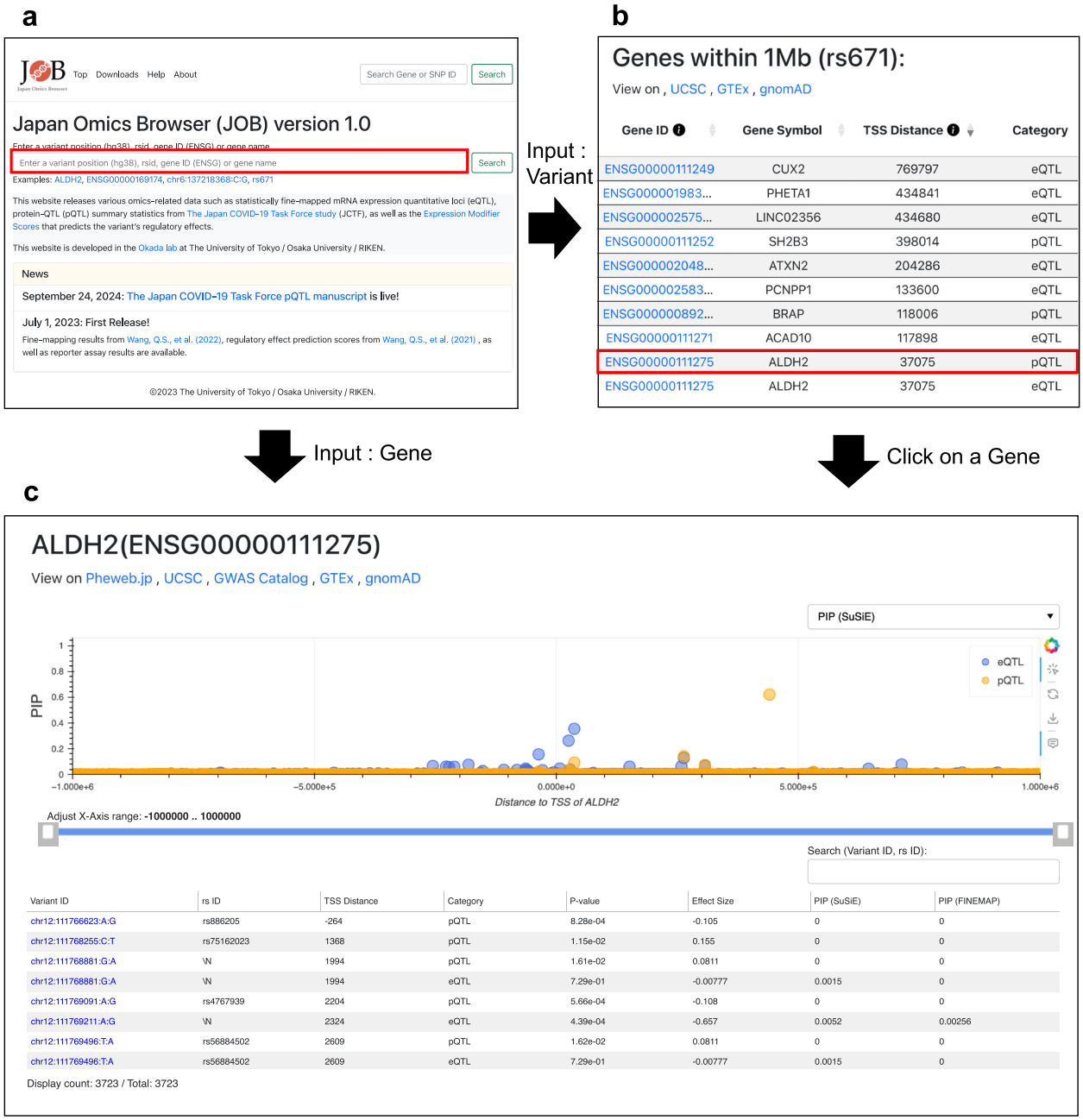

Fig. 2 User interface of JOB. **a** Users enter the variants and genes of interest on the search page. **b** All genes with a transcription start site (TSS) within ± 1 Mb of the variant position are displayed. **c** All variants within ± 1 Mb of the gene's TSS are displayed. Detailed information of the variants plotted on the figure is summarized in a table

2. The same features for protein expression quantitative loci (pQTL).
3. PIPs from statistical fine-mapping of UK Biobank phenotype in SuSiE.
4. Machine learning-based prediction of gene regulatory effect variants for 49 GTEx tissues.
5. The allelic effect (log2 fold change) measured by the massively parallel reporter assay (MPRA).

The variant page also includes additional useful features, such as hyperlink to external platforms (e.g. UCSC

and GTEx), as well as redirection to the gene page by clicking on the ensembl ID within the table.

When an ensembl gene ID or canonical gene name is entered on the search page, users will be redirected to the gene page. On the gene page, variants within a cis-window (i.e. within 1 Mb from the TSS) are displayed, allowing users to qualitatively assess functional variants in the cis-regulatory region (Fig. 2c). For a more detailed overview of the gene page layout, please see Supplementary Material 1: Fig. S2. The first section of the page displays the posterior inclusion probability (PIP) obtained from statistical fine-mapping of e/pQTLs from Japan COVID-19 Taskforce [20]. The second section shows PIP obtained from statistical fine-mapping of 94 complex traits from UK Biobank (UKBB [25]). The third section shows the expression modifier score (EMS [22]) for gene regulation, a score that provides quantitative estimation of per-tissue gene regulatory activity across 49 tissues from GTEx [13]. The fourth section presents experimental validation score from the massively parallel reporter assay (MPRA [26]) for two selected cell lines, K562 and HepG2, that are available for >10,000 variants ascertained for high-PIP and/or significant association with complex traits in UKB or other large biobank studies [12]. Users can freely customize the displayed content using the dropdown menu in the top right corner of each graph. For example, displaying the PIP instead of the *p*-value is useful for identifying the causal variant of a gene (Supplementary Material 1: Fig. S3). Additionally, the bottom slider allows for easy zooming in on specific regions of interest. Hovering the cursor over a variant of interest on the graph reveals detailed information about its score. When exploring the gene page, users will discover numerous variants showing peaks near the TSS. However, it is important to note that there are also relevant variants present at distant locations. At the bottom of the figure, a comprehensive table displays the scores, and clicking on a variant of interest provides further information about its impact on other gene expressions on the variant page (Fig. 2c). Users can also cross-reference gene information using other platforms such as the GWAS catalog for the searched gene [27].

Integrative analysis of eQTL, pQTL, and reporter assay data

We summarize a set of examples demonstrating the utility of JOB. The First example is rs74953707 (chr1:155,250,599:G:A in hg38), where all the (1) regulatory activity predicted from thousands of functional annotations (EMS), (2) putative causal regulatory evidence from statistical fine-mapping of eQTL signal in the JCTF cohort, and (3) transcriptional regulatory activity measured in high-throughput reporter assay (MPRA) colocalize to clearly highlight the regulatory effect of the variant on gene *FAM189B* (posterior inclusion probability: PIP = 1.0) (Fig. 3). The variant is likely to increase the expression of *FAM189B* in the whole blood (beta = 0.079 in eQTL analysis, and log2 fold change = 1.07 in MPRA, both pointing to the same effect direction). Notably, this variant is rare in European (MAF ~ = 0.003 in gnomAD) and the association has not been reported in other major database, highlighting the value of JOB not only as a user-friendly integrative browser but also as a large statistical and functional fine-mapping database of East Asian studies.

The second example is rs35083095 (chr3:4,362,134:T:C in hg38), which shows high pQTL signal and putatively regulates genes *SETMAR* (PIP = 0.78; beta = −0.105) and *SUMF1* (PIP = 0.73; beta = −0.157) through changes in protein expression levels in the plasma. Although the mRNA expression regulatory effect of the variant on these two genes are previously suggested through large scale eQTL studies [13], the penetrance of the regulatory effect to the downstream of central dogma has not been confirmed in detail with a browser-level visualization to date. JOB allows visual and numerical confirmation of the eQTL effected penetrated to the protein expression regulation level with consistent effect sizes (both eQTL and pQTL effects are negative), providing further characterization of its previously known eQTL effect (Fig. 4).

Variant-gene prioritization across UKBB phenotype

As a case study in interpreting a complex trait-associated locus using JOB, we walk through the process of exploring the regulatory effect of rs10411704 on *CD22* expression as well as its downstream effects on complex traits. Our analysis reveals a putatively causal association between rs10411704 and *CD22* expression levels (50 < fold enrichment of EMS and pQTL PIP = 0.37). Notably, this regulatory effect is found to co-localize with hematopoietic traits (PIP = 0.28) within the UK Biobank dataset (Fig. 5). The observed regulatory effect of rs10411704 on *CD22* expression suggests its involvement in modulating the function of B cells and potentially influencing immune responses. CD22, a well-known B-cell specific antigen [28], plays a crucial role in the regulation of B-cell development, activation, and tolerance. Dysregulation of *CD22* has been implicated in various hematological disorders and autoimmune diseases [29]. Although direct functional validations are warranted, our analysis of the functional annotation of the variant based on sequence context suggests that the variant disrupts the *SETDB1* binding motif (e-value < 1e-300 in https://www.factorbook.org). SETDB1 is an H3K9 methyltransferase involved in chromatin compaction and gene expression regulation. It plays a crucial role in the differentiation and proliferation of B cells and other immune cells [30].

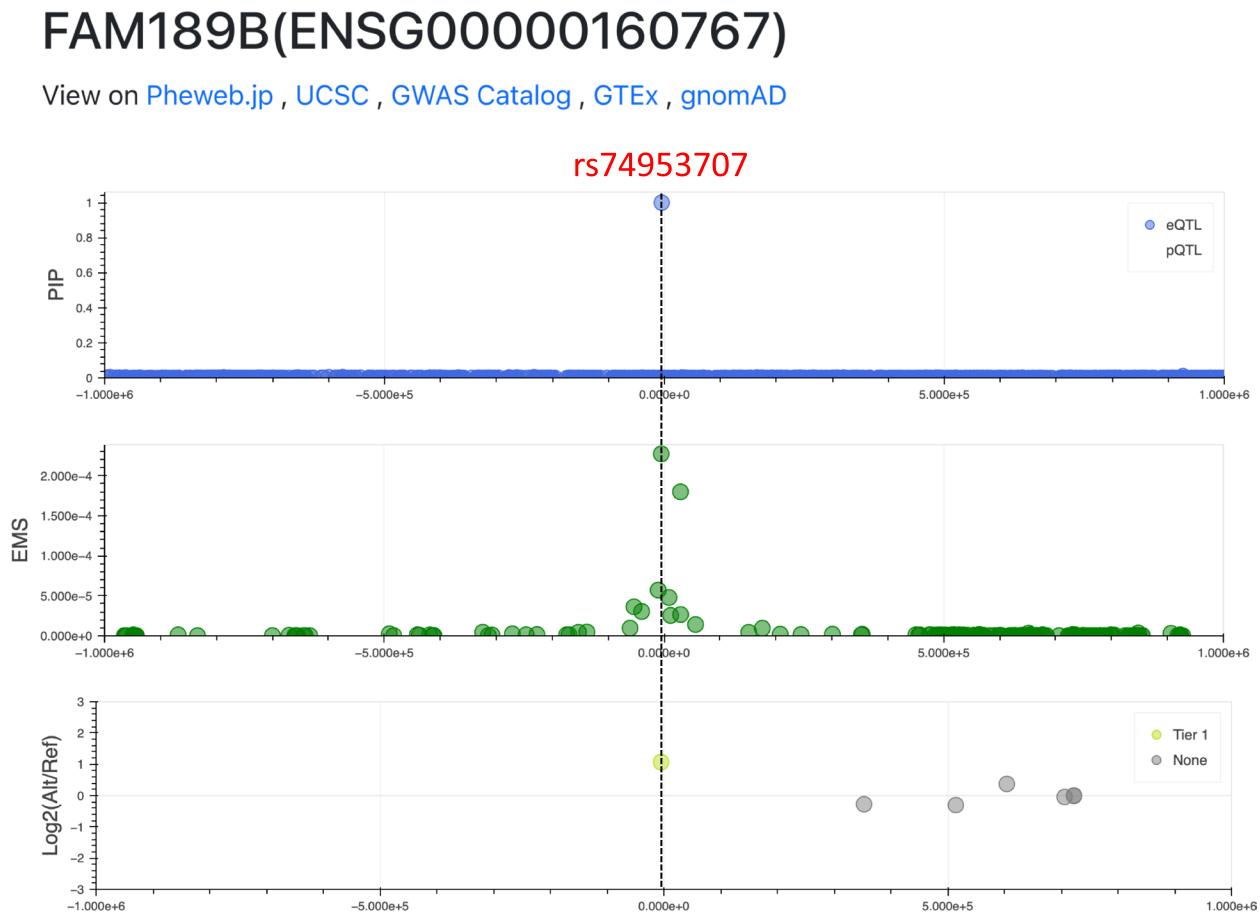

Fig. 3 An example of putative causal eQTL prioritized by EMS and MPRA. The example of rs74953707, where all the EMS, MPRA, and eQTL signal colocalize to clearly highlight the regulatory effect on gene FAM189B (posterior inclusion probability: PIP=1.0). Notably, this variant is rare (MAF~=0.003) in European and the association has not been reported in other major database, highlighting the value of JOB

These findings provide valuable insights into the genetic mechanisms underlying *CD22* expression variation and its putative causal contributions to hematopoietic traits. The integration of multi-omics data within JOB allows for a comprehensive analysis of variant-gene associations and their potential molecular implications in complex traits and diseases.

Discussion

With the advancement of high-throughput omics technologies, more and more large-scale genomic datasets are generated. Variant annotation from large populations and integrative omics analysis are essential for functional interpretation and downstream functional analysis of variants. However, analyzing these data from various data resources can be practically and computationally challenging, serving as a bottleneck for evaluating the impact of genetic variations.

In this manuscript, we presented the Japan Omics Browser (JOB), aiming to provide widely and easily accessible multi-omics data, enabling comprehensive and integrative analysis of omics data from the Japanese population. Using JOB, users can qualitatively understand the regulatory effects of individual variants and analyze potential downstream consequences. The system infrastructure is hosted on the Google Cloud Platform, ensuring a robust and persistent service. To date, JOB is one of the few platforms that integrates all the layers of multi-omics information in a single page, and is unique in that it is focused on East Asian population.

We showed how JOB can be utilized in multiple scenarios, each presenting its unique value. First, taking rs74953707 near *FAM189B* gene as an example, we showed that JOB allows discoveries of regulatory variants that are rare in European populations, leveraging multiple layers of evidence such as fine-mapping results in Japanese population, functionally predicted regulatory activity score (EMS) and direct experimental validation in MPRA. Second, focusing on rs3508309 near *SUMF1* and *SETMAR*, we showed how JOB enables a deeper

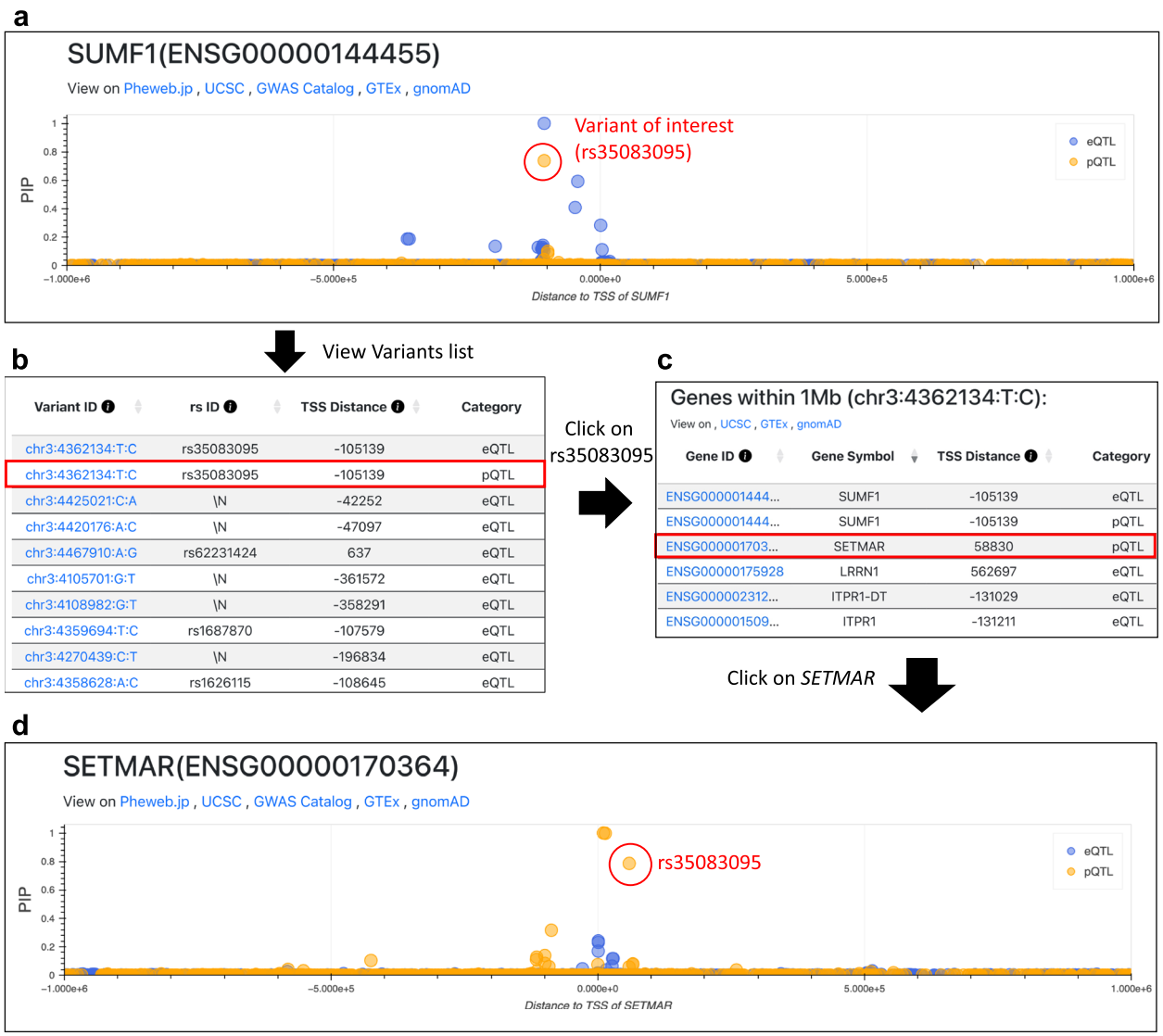

Fig. 4 An example of variant-gene search. **a** In the example of an e/pQTL plot for *SUMF1*, it can be observed that rs35083095 shows a high pQTL causal probability (PIP). **b** Searching and selecting rs35083095 from the variants list. **c** Displaying the list of genes within the cis-window of the variant, and selecting *SETMAR*. **d** In the e/pQTL plot for SETMAR, it also shows a high pQTL PIP for rs35083095. This demonstrates the significant impact of rs35083095 on protein expression of *SUMF1* and *SETMAR*

understanding of regulatory mechanisms by incorporating both eQTL and pQTL data. Third, the example of rs10411704 upstream of *CD22* gene highlighted JOB's potential to connect variant's regulatory effect to complex traits. Finally, all these individual examples are compelling showcases of JOB's user-friendly interface enabling visual confirmation of putative variants directly in the browser, without complex software.

We note that, although integrated visualization of multi-omics information in JOB is effective in variant interpretation, a careful evaluation such as statistical colocalization analysis [32], external data comparison, and experimental validation is essential for elucidating the effect of genetic variations. Although our MPRA information covers across the genome, the number of the MPRA-examined variant, at the order of tens of thousands, is a small fraction of the human genome variation.

Future updates include, but are not limited to, increased coverage of MPRA data, microRNA expression QTLs [33] and single-cell RNA-seq data [34], all further aiding functional interpretation of trait-associated variants at single variant resolution. For example, blood cell type-specific eQTL summary statistics, called from uniformly processed single-cell RNA-seq data from over

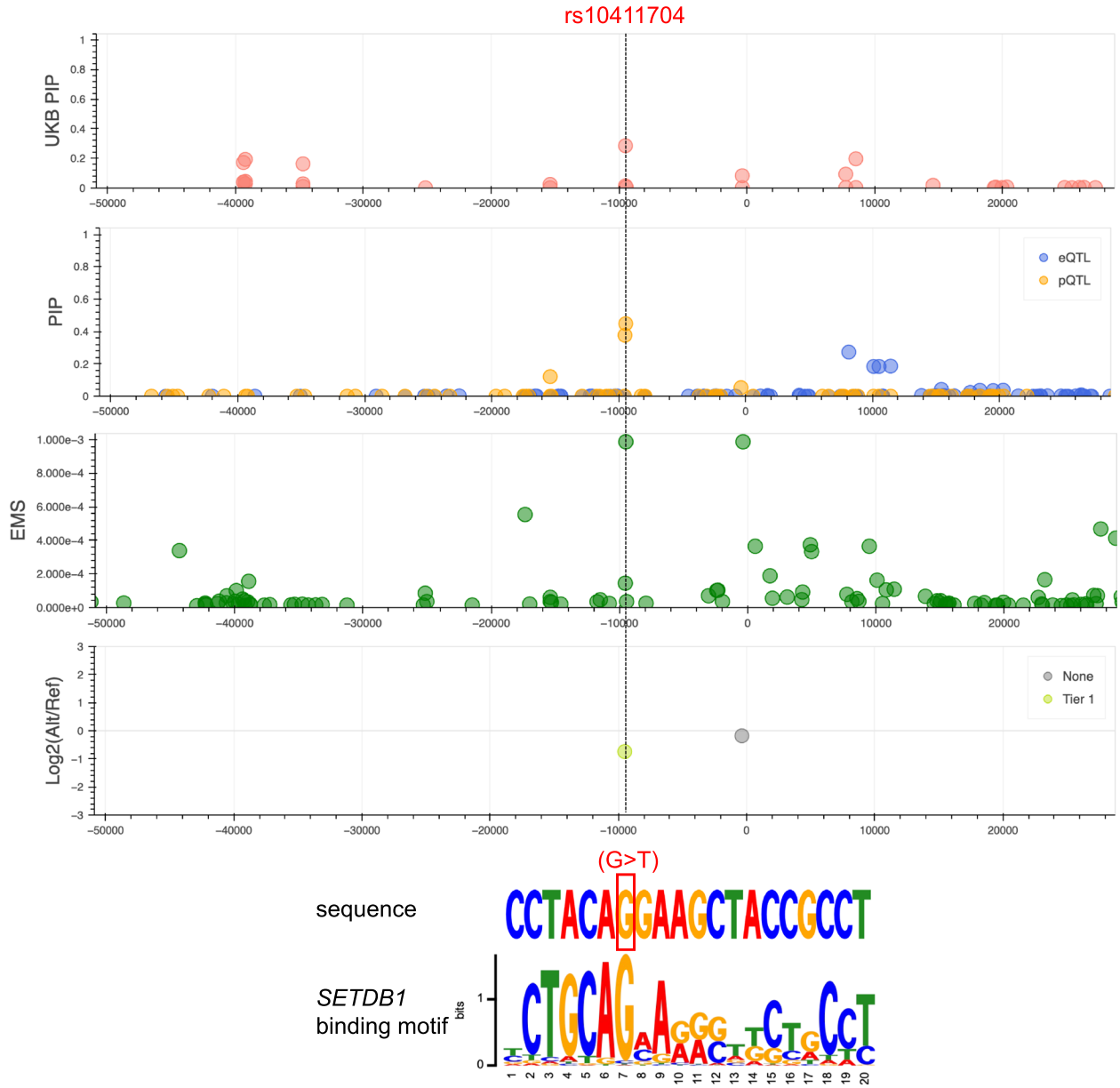

Fig. 5 An example of interpreting complex trait-associated locus. An example of rs10411704, an upstream variant of CD22. From top to bottom, it shows the PIP ofhematopoietic traits in the UK Biobank, pQTL PIP, EMS in Whole Blood, and MPRA in K562 cell line.rs10411704 exhibits high scores in all measures and has been experimentally validated as a variantassociated with gene expression. Additionally, the genetic variation in the sequence may lead to areduction in the binding motif of SETDB1, potentially regulating the expression of CD22. The sequencelogo was generated using Factorbook (https://www.factorbook.org/) [31] and the binding motif is in https://www.factorbook.org/tf/human/SETDB1/motif/ENCSR000EWI

1,500,000 peripheral blood mononuclear cells (PBMCs) in over 200 Japanese individuals, is planned to be added in the future release within 2025. With public accessibility, JOB empowers comprehensive omics analyses, facilitating investigations into gene expression's relevance to complex traits and diseases, providing significant value to the research community.

Conclusion

In conclusion, the Japan Omics Browser (JOB) provides a comprehensive platform for visualizing regulatory effects of variants on nearby genes by integrating omics data. It offers visualizations of various layers of regulatory evidence at mRNA and protein levels, enabling the exploration of disease mechanisms. The browser includes

regulatory variants validated by in-house reporter assays and will be continuously updated with additional omics data, enhancing functional interpretation of trait-associated variants. The presented examples highlight the effectiveness of JOB in identifying regulatory effects and uncovering novel association between genetic variations and diseases. JOB is publicly and freely available at https://japan-omics.jp/.

Methods

JOB architecture

The Japan Omics Browser (JOB) is an integrated web application hosted on the Google Cloud Platform. It utilizes Google App Engine as the web server and Google Cloud SQL as the database server, ensuring scalability, reliability, and security while managing database operations and executing queries [35–37]. SSL is enabled by default in App Engine, ensuring that HTTPS connections automatically encrypt communications between the server and client. The system is designed to be flexible and scalable, allowing for future database expansion. The backend is implemented using Flask (https://github.com/pallets/flask/), a microframework for web applications built with Python, to handle request processing and interact with the database. The frontend, based on HTML/CSS, incorporates Bootstrap (https://github.com/twbs/bootstrap) and JQuery (https://github.com/jquery/jquery) to achieve responsive design and interactive elements. Bokeh (https://github.com/bokeh/bokeh) is used to plot the data. The JOB source code is available on GitHub repository (https://github.com/ytakahashi-statgen/Japan-Omics-Browser).

Fine-mapped eQTLs and pQTLs in a Japanese population

JOB utilized statistically fine-mapped mRNA expression quantitative loci (eQTL, $n=$ 1,019 samples) and protein-QTL (pQTL, $n=1{,}384$ samples) summary statistics obtained from the Japan COVID-19 Task Force [20]. Japan COVID-19 Task Force is a nationwide multicentre consortium, collecting genotype, disease severity, and multiple omics information from COVID-19 patients. The study followed a pipeline based on the GTEx pipeline for eQTL calling and utilized tools such as STAR and RSEM for mRNA quantification. Since the eQTL and pQTL fine-mapping were performed using FINEMAP [24] and SuSiE [23] methods, posterior inclusion probabilities (PIPs) from both algorithms are available in JOB. In the study, protein expression was measured using Olink Explore 3072 platform, and identification of pQTL was performed on the normalized protein expression (NPX) scale. When protein expression data is missing for a gene (which is the case for the majority of genes, since Olink Explore 3072 covers only near 3,000 genes), the pQTL PIPs are unavailable (although other omics data are available). Further details of the summary statistics generation are available in [20].

Fine-mapping of UKBB traits

JOB shows fine-mapping results for 94 heritable traits in the UK Biobank, performed in Kanai et al. [38] and available in public. We show the effect size Beta and SE(Beta) values as obtained from summary statistics (i.e. the effect sizes are marginal effect sizes, instead of posterior). Detailed methodology as well as the source data can be found in https://www.finucanelab.org/data.

Machine-learning based per-tissue regulatory effect prediction score (EMS)

The expression modifier score (EMS) is a score predicting the regulatory effect of individual variant on nearby gene expressions in various tissues, utilizing thousands of functional genomics features. As described in [22], EMS is a score from a random forest predictor trained to predict the result of fine-mapping of GTEx v8 eQTL data from the FINEMAP [24] and SuSiE [23] methods, where variant-gene pairs were labeled as positive or negative based on the posterior inclusion probability (PIP) thresholds, and is scaled to have a probability unit. Functional genomics features including TSS distance, LDSC baseline model [39], features from Roadmap Epigenomics Consortium [40], and deep-neural network derived activity scores (Basenji [41]) were annotated for each variant-gene pair. As described in detail in [22], EMS was evaluated to have a high performance as measured in the Area under receiver operating characteristic (AUROC) and Area under precision-recall curve (AUPRC) in held-out fine-mapped eQTLs within GTEx, as well as enrichment in reporter-assay validated regulatory variants [42]. We obtained EMS that was publicly available for the majority of GTEx tissues ($n=$ 49 tissues) and made it visible in JOB.

Massive parallel reporter assay (MPRA)

As part of the Japan COVID-19 Task Force study [20], MPRA was conducted to validate a subset of fine-mapped eQTLs. As described in [26], the MPRA employed a library of 24,000 oligos, enabling the evaluation of approximately 12,000 variants across the genome, predominantly focusing on medium to high PIP variants in their eQTL fine-mapping. MPRA experiments were performed in HepG2 and K562 cells, where the data for both cell types are available in JOB. Variants passing the false discovery rate (FDR) thresholds were classified as tier 1 or tier 2 expression-modifying variants (emvars; FDR <0.01 for Tier 1 and FDR <0.1 for Tier 2).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11639-1.

Supplementary Material 1

Acknowledgements
We would like to sincerely thank all the participants involved in this study, and all the members of Japan COVID-19 Task Force for their support. Y.T is supported by the Osaka University Medical Doctor Scientist Training Program.

Authors' contributions
Y.T. and Q.S.W. designed the study. Y.T. implemented the browser with input from Q.S.W. and Y.O. Y.T. wrote the manuscript with input from Q.S.W. and Y.O. T.H., H.N., F.I., K.F., S.I., and S.M. provided intellectual input throughout the study, offered comments, and helped edit the manuscript. Q.S.W. and Y.O. supervised the work. All authors and the JCTF contributed to the generation of the primary data incorporated in the study, provided inputs and approved the manuscript.

Funding
This study was supported by AMED (JP23kk0305022, JP22ek0410075, JP23 km0405211, JP23 km0405217, JP23ek0109594, JP23ek0410113, JP223fa627002, JP223fa627010, JP233fa627011, JP23zf0127008, JP24tm0524002, JP24wm0625504, JP24gm1810011), JST FOREST (JPMJFR225Y), JST Moonshot R&D (JPMJMS2021, JPMJMS2024), JSPS KAKENHI (25H01057, 23 K14233), the Nakajima Foundation, the Uehara Memorial Foundation, Takeda Science Foundation, Ono Pharmaceutical Foundation for Oncology, Immunology, and Neurology, Bioinformatics Initiative of Osaka University Graduate School of Medicine, Institute for Open and Transdisciplinary Research Initiatives, Center for Infectious Disease Education and Research (CiDER), and Center for Advanced Modality and DDS (CAMaD), Osaka University, RIKEN TRIP initiative (AGIS).

Data availability
All data is freely available from https://japan-omics.jp/. The source code for JOB is available at https://github.com/ytakahashi-statgen/Japan-Omics-Browser. The summary statistics of QTL analyses are available at the National Bioscience Database Center (NBDC) Human Database (accession code: hum0343; https://humandbs.dbcls.jp/en/). The expression modifier score (EMS) for 49 tissues are available at https://www.finucanelab.org/data. The UKBB fine-mapping results are deposited at https://www.finucanelab.org/data.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
Q.S.W. is an employee of Calico Life Sciences. The other authors declare no competing interests.

Received: 22 October 2024 Accepted: 25 April 2025
Published online: 07 May 2025

References

1. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
2. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.
3. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10.
4. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. Elife. 2017;6:e27041.
5. Zhang J, Dutta D, Köttgen A, Tin A, Schlosser P, Grams ME, et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. Nat Genet. 2022;54:593–602.
6. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558:73–9.
7. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017;18:83.
8. Forny P, Bonilla X, Lamparter D, Shao W, Plessl T, Frei C, et al. Integrated multi-omics reveals anaplerotic rewiring in methylmalonyl-CoA mutase deficiency. Nat Metab. 2023;5:80–95.
9. Liu Z, Zhao Y, Kong P, Liu Y, Huang J, Xu E, et al. Integrated multi-omics profiling yields a clinically relevant molecular classification for esophageal squamous cell carcinoma. Cancer Cell. 2023;41:181-195.e9.
10. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.
11. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12:996–1006.
12. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet. 2021;53:1415–24.
13. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369:1318–30.
14. Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods. 2021;18:1196–203.
15. Chen KM, Wong AK, Troyanskaya OG, Zhou J. A sequence-based global map of regulatory activity for deciphering human genetics. Nat Genet. 2022;54:940–9.
16. Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res. 2021;49:D1311–20.
17. Wu D, Dou J, Chai X, Bellis C, Wilm A, Shih CC, et al. Large-Scale Whole-Genome Sequencing of Three Diverse Asian Populations in Singapore. Cell. 2019;179:736-749.e15.
18. Feng YCA, Chen CY, Chen TT, Kuo PH, Hsu YH, Yang HI, et al. Taiwan Biobank: A rich biomedical research database of the Taiwanese population. Cell Genom. 2022;2:100197.
19. Buniello A, Suveges D, Cruz-Castillo C, Llinares MB, Cornu H, Lopez I, et al. Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. Nucleic Acids Res. 2025;53:D1467–75.
20. Wang QS, Edahiro R, Namkoong H, Hasegawa T, Shirai Y, Sonehara K, et al. The whole blood transcriptional regulation landscape in 465 COVID-19 infected samples from Japan COVID-19 Task Force. Nat Commun. 2022;13:4830.
21. Wang QS, Hasegawa T, Namkoong H, Saiki R, Edahiro R, Sonehara K, et al. Statistically and functionally fine-mapped blood eQTLs and pQTLs from 1,405 humans reveal distinct regulation patterns and disease relevance. Nat Genet. 2024;56:1–14.
22. Wang QS, Kelley DR, Ulirsch J, Kanai M, Sadhuka S, Cui R, et al. Leveraging supervised learning for functionally informed fine-mapping of cis-eQTLs identifies an additional 20,913 putative causal eQTLs. Nat Commun. 2021;12:3394.
23. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. J R Stat Soc Series B Stat Methodol. 2020;82:1273–300.
24. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32:1493–501.
25. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9.

26. Klein JC, Agarwal V, Inoue F, Keith A, Martin B, Kircher M, et al. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. Nat Methods. 2020;17:1083–91.
27. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51:D977–85.
28. Sgroi D, Varki A, Braesch-Andersen S, Stamenkovic I. CD22, a B cell-specific immunoglobulin superfamily member, is a sialic acid-binding lectin. J Biol Chem. 1993;268:7011–8.
29. Macauley MS, Crocker PR, Paulson JC. Siglec-mediated regulation of immune cell function in disease. Nat Rev Immunol. 2014;14:653–66.
30. Johnson E, Salari K, Yang S. SETDB1: A perspective into immune cell function and cancer immunotherapy. Immunology. 2023;169:3–12.
31. Pratt HE, Andrews GR, Phalke N, Purcaro MJ, van der Velde A, Moore JE, et al. Factorbook: an updated catalog of transcription factor motifs and candidate regulatory motif sites. Nucleic Acids Res. 2022;50:D141–9.
32. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. 2014;10:e1004383.
33. Sonehara K, Sakaue S, Maeda Y, Hirata J, Kishikawa T, Yamamoto K, et al. Genetic architecture of microRNA expression and its link to complex diseases in the Japanese population. Hum Mol Genet. 2022;31:1806–20.
34. Edahiro R, Shirai Y, Takeshima Y, Sakakibara S, Yamaguchi Y, Murakami T, et al. Single-cell analyses and host genetics highlight the role of innate immune cells in COVID-19 severity. Nat Genet. 2023;55:753–67.
35. Google App Engine. Google Cloud. [cited 2024 Apr 9]. Available from: https://cloud.google.com/appengine.
36. Google Cloud SQL. Google Cloud. [cited 2024 Apr 9]. Available from: https://cloud.google.com/sql.
37. Google Cloud Storage. Google Cloud. [cited 2024 Apr 9]. Available from: https://cloud.google.com/storage.
38. Kanai M, Ulirsch JC, Karjalainen J, Kurki M, Karczewski KJ, Fauman E, et al. Insights from complex trait fine mapping across diverse populations. medRxiv. 2021. Available from: https://doi.org/10.1101/2021.09.03.21262975.
39. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47:1228–35.
40. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30.
41. Kelley DR, Reshef YA, Bileschi M, Belanger D, McLean CY, Snoek J. Sequential regulatory activity prediction across chromosomes with convolutional neural networks. Genome Res. 2018;28:739–50.
42. van Arensbergen J, Pagie L, FitzPatrick VD, de Haas M, Baltissen MP, Comoglio F, et al. High-throughput identification of human SNPs affecting regulatory element activity. Nat Genet. 2019;51:1160–9.

Publisher's Note